International Journal of Advanced Statistics and Probability, 13 (1) (2026) 27-31

SPC

International Journal of Advanced Statistics and Probability

Website: www.sciencepubco.com/index.php/IJASP
https://doi.org/10.14419/9n9x7b95
Research paper

A Note on Some Hidden Properties of The Variance And Other Closely Related Measures

Mohammad Fraiwan Al-Saleh *, Hashim Abdallah Jarrah

Department of Statistics, Yarmouk University-Jordan
**Corresponding author E-mail: m-saleh@yu.edu.jo*

Received: December 27, 2025, Accepted: February 5, 2026, Published: February 10, 2026

Abstract

Several important hidden properties about population variance and sample variance are highlighted. Very interesting, important queries are raised about the definition of the mean absolute deviation, the sample variance, and standard deviation. In. In addition, the close relationship between variance and the coefficient of variation is emphasized and highlighted. It is noted that these two measures of variation are functions of the same two quantities, the average of the squares and the square of the average of the data. A very important comment is raised about the definition of the mean absolute deviation.

Keywords: *Population Variance; Sample Variance; Mean Absolute Deviation; Coefficient of Variation; Standard Deviation; Stratified Random Sampling.*

1. Introduction

Statistics is the science of making inferences about the population characteristics using information in a suitable representative sample. Frequently, some very important statistical facts and concepts are never mentioned in classrooms or mentioned but without enough details. Giving more information about them motivates students toward learning. In this paper, details about some of the hidden properties of the variance and other closely related measures are highlighted. Some of these properties may be found in the contents of some published papers, seminars, or conference talks, and others never discussed.

The literature is full of papers that contain very important and interesting comments about some statistical concepts. The reasons for the $(n-1)$-Divisor in the definition of sample variance was closely discussed by Al-Saleh [1]. Al-Saleh [2] highlighted some interesting information about the Poisson distribution using the fact that the mean and variance of this distribution are equal. The relation between the standard deviation and the mean absolute deviation was discussed by Al-Saleh and Yousif [3]. Casella and Berger [4] looked in their book at some properties of the sample variance and showed that it can be obtained in more than one way. Sen's Inequality, which is an inequality that gives an upper bound of the probability that the sum of non-negative random variables is greater than a given level, was considered by Balakrishnan and Balasubramanian [5]. Heien and Rahman [6] discussed the randomness of the Digits of π. Very strange properties of the Cauchy Distribution were given by Al-Saleh and Maabreh [7]. Al-Saleh and Jarrah [8] highlighted some properties of some of the main sampling techniques. Properties of the arithmetic, geometric, and harmonic means were discussed by Heinbockel [9]. A note on the generating function of a negative hypergeometric was given by Khan [10]. A note on the estimation of the location parameter of the Cauchy distribution was given by Bloch[11]. The unimodality of the likelihood for this strange distribution was discussed by Copas [12]. A note on estimation based on a random sample from a Cauchy distribution was given by Rothenberg et al. [13].

Statistical teachers usually tell students that the reason for the division by (n-1) in the formula of the sample variance. S^2 is to make it an unbiased estimator of the population variance σ^2 without mentioning that this is true only when the population is infinite. For a finite population of size N, $\frac{N-1}{N}S^2$ is an unbiased estimator of σ^2.

More work on statistical pedagogy or robustness of dispersion measures can be seen in Azmay et al. [14] and Schoen et al. [15]

A close look at the population variance, sample variance, and some other closely related measures is the content of Section 2. Very important queries are raised about the definition of the mean absolute deviation, the sample variance, and the standard deviation. Conclusions are given in section 3.

2. Hidden Properties of The Variance and Other Related Measures

Measures of central tendency and dispersion play a central role in the summary of numerical data. While the mean is widely understood and intuitively interpreted, variance is often viewed as a more abstract quantity. Even among well-trained students, several of its fundamental properties remain unclear or underappreciated. This section explores a number of these properties and examines how variance is connected to other measures of variability, such as the mean absolute deviation(MAD) and the coefficient of variation(CV).

Population Variance

Suppose that the population consists of N numerical values of the variable of interest. U ,$u_1, u_2, \ldots, u_N$, then the population mean μ and variance σ^2 are:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} u_i,\ \sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(u_i - \mu)^2 = \frac{1}{N}\sum_{i=1}^{N} u_i^2 - \mu^2$$

Clearly, $\sigma^2$is the average of the squared distances of the observations from their mean value. This interpretation highlights the role of the mean as a reference point(central point) and explains why squaring the deviations is essential in the definition. Several useful properties of the variance follow directly from this formulation and are discussed below:

1) $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(u_i - \mu)^2 = \frac{1}{N}\min_a\left(\sum_{i=1}^{N}(u_i - a)^2\right)$.

Thus, the minimum value of $\sum_{i=1}^{N}(u_i - a)^2$occurs when $a = \mu$. In other words, $\mu$is the closest reference point to the data.
For any real number$a$, $\frac{d}{da}\left(\sum_{i=1}^{n}(u_i - a)^2\right) = -2N(\mu - a) \overset{set}{=} 0 \Rightarrow a = \mu$. The 2nd derivative is positive. Therefore,$min_a\left(\sum_{i=1}^{N}(u_i - a)^2\right) = \sum_{i=1}^{N}(u_i - \mu)^2 = N\sigma^2$. Thus, σ^2 is the smallest of all "average squared distances" of the observed values of $U$from any real point". In other words, μ is the closest point to the population values; μ is the point that makes the value of "the average squared distance" the smallest.

2) $\sum_{i=1}^{N}(u_i - \mu) = \sum_{i=1}^{N} u_i - N\mu = 0$; that is one of the reasons for squaring $(u_i - \mu)$ in the formula of σ^2.

3) If $u_{(1)}, u_{(2)} \ldots, u_{(N)}$ Are the values in ascending order, then $\mu = \sum_{i=1}^{N} u_{(i)}/N$; Thus, $u_{(1)} \le \mu \le u_{(N)}$. Also, $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(u_i - \mu)^2 = \frac{1}{N}\sum_{i=1}^{N} u_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N} u_i\right)^2$.

Thus, σ^2 Is the difference between "Average of the Squares" & "Square of the Average".
This can be another definition of σ^2. Clearly, squaring first and taking the average is larger than taking the average first and then squaring(using Jenson's Inequality). Thus, the variance is non- negative. It is zero if all values of U are the same; the variable has only one value (constant).

4) To decide about how large or small a quantity is, it is very important to know its extreme values. Now, what is the maximum value of the variance?

Fact: If $a \le u_i \le b,\ for\ i = 1, \ldots, N$, then $0 \le \sigma^2 \le (b - a)^2/4$.
For even N, the maximum occurs when $N/2$of the values equal to the smallest possible value of U, which is a, and each element of the other half equals the maximum possible value, which is b. In this case, we have,

$$Max(\sigma^2) = \frac{1}{N}\left[\frac{N}{2}\left(a - \frac{1}{2}(a + b)\right)^2 + \frac{N}{2}\left(b - \frac{1}{2}(a + b)\right)^2\right] = \frac{1}{4}(b - a)^2.$$

If $N$is odd, then it can be shown easily that the absolute maximum occurs when$\frac{1}{2}(N - 1)$ of the values$= a\ \&\ \frac{1}{2}(N + 1) = b$or $\frac{1}{2}(N - 1) = b\ \&\ \frac{1}{2}(N + 1) = a$. Thus,

$$\sigma^2 \le \frac{(N^2-1)(b-a)^2}{4N^2} \le \frac{1}{4}(b - a)^2.$$

In other words,

$$\sigma^2 \le \frac{\left(u_{(N)} - u_{(1)}\right)^2}{4},\ \sigma \le \frac{1}{2}\left(u_{(N)} - u_{(1)}\right) = \frac{1}{2}Range.$$

This inequality is known, but probably rarely mentioned in classrooms. Thus, maximum variability in the population occurs when the population consists of two homogeneous groups of equal size. This can be seen when we deal with the binomial population.
For a binomial random variable, $X \sim Bin(n, p)$, we have$\sigma^2 = np(1 - p)$, with a maximum value of $\sigma^2 = \frac{n}{4}$ occurs when $p = 0.5$, which means that the maximum occurs when half of the values of n Bernoulli trials are ones and a half of them are zeros.

5) $\sigma^2 = \sum_{i=1}^{N}(u_i - \mu)^2/N = \sum_{j,i=1,i\ne j}^{N}(u_i - u_j)^2/2N^2 = \sum_{j>i=1}^{N}(u_i - u_j)^2/N^2$.

Proof: $\frac{1}{2N^2}\sum_{j=1}^{N}\sum_{i=1}^{N}(u_i - u_j)^2 = \frac{1}{2N^2}\sum_{j=1}^{N}\sum_{i=1}^{N}((u_i - \mu) + (\mu - u_j))^2 = \frac{1}{2N^2}\sum_{j=1}^{N}\left[\sum_{i=1}^{N}(u_i - \mu)^2 + 2N(\mu - u_j)\sum_{i=1}^{N}(u_i - \mu) + N(u_j - \mu)^2\right]$

$$= \frac{1}{2N^2}\left[N\sum_{i=1}^{N}(u_i - \mu)^2 + N\sum_{j=1}^{N}(u_j - \mu)^2\right]$$

$$= \frac{1}{2N}\left[\sum_{i=1}^{N}(u_i - \mu)^2 + \sum_{j=1}^{N}(u_j - \mu)^2\right] = \frac{1}{N}\left[\sum_{i=1}^{N}(u_i - \mu)^2\right] = \sigma^2.$$

Thus, it can be seen explicitly that. σ^2 is also a measure of the variations among the elements of the population.
The Coefficient of Variation (CV)
$CV$is another measure of variability:$CV = \frac{\sigma}{|\mu|}; \mu \ne 0$.

$CV^2 = \frac{\sigma^2}{\mu^2} = \frac{\left(\frac{1}{N}\sum_{i=1}^{N} u_i^2\right)-\left(\frac{1}{N}\sum_{i=1}^{N} u_i\right)^2}{\left(\frac{1}{N}\sum_{i=1}^{N} u_i\right)^2} = \frac{\left(\frac{1}{N}\sum_{i=1}^{N} u_i^2\right)}{\left(\frac{1}{N}\sum_{i=1}^{N} u_i\right)^2} - 1$.

Thus, $CV$and $\sigma^2$are functions of the same two quantities: "Average of the Squares " & "Square of the Average"; the two quantities are compared based on their difference or ratio. $\sigma^2$is a measure that is obtained based on the difference between the two quantities, while $CV^2$is a measure that is obtained based on the ratio of the two quantities. Clearlyσ has the same unit of measurement as the variable u. This makes it an unsuitable measure to use to compare data sets with different units of measurement (for example: temperature in $F°$ or in $C°$, weight in kg or pounds, etc.). CV is free of units, so it can be used to compare two data sets regardless of their unit of measurement. For example, the standard deviation of the data 1, 2, 3 is 1, and the $CV = 0.5$; the standard deviation of the data (99, 100, 101) is 1 and $CV = 0.01$. But clearly, the variability in the first set is much higher than in the second set. Thus, Variance is not an appropriate measure for comparing the variability between two groups; CV is a suitable measure to compare the variability between two groups.

Mean Absolute Deviation(MAD)
The mean absolute deviation of $u_1, u_2, \ldots, u_N$is

$$MAD = \sum_{i=1}^{N} |u_i - \mu| \,/N.$$

MAD provides an alternative measure of dispersion by averaging the absolute distances of observations from a reference point. Unlike variance, it preserves the original unit of measurement of the data. A natural question arises regarding the choice of the reference point used in its definition. While deviations from the mean are commonly considered, it can be shown that the median minimizes the sum of absolute deviations. This distinction highlights an important conceptual difference between measures based on squared deviations and those based on absolute deviations. Now,

$min_a\left(\frac{1}{N}\sum_{i=1}^{N}|u_i - a|\right) = \frac{1}{N}\ \sum_{i=1}^{N}|u_i - Median|$ It is not $\sum_{i=1}^{N}|u_i - \mu| \,/N$.

Thus, why $MAD = \frac{1}{N}\sum_{i=1}^{n}|u_i - \mu|$, not $\frac{1}{N}\sum_{i=1}^{N}|u_i - Median|$?
This is a very important question that needs to be discussed by statisticians.
$\sum_{i=1}^{N}(u_i - a) = 0$ only if $a = \mu$. Thus, total variability can be measured using. $|u_i - \mu|$ or $(u_i - \mu)^2$.
The following are some extra comments about σ and MAD:

1) The unit of σ^2 is the square of the unit of the data, while the unit of σ And MAD is the same as the unit of the data.

2) $\left[\sum_{i=1}^{N}(u_i - \mu)^2\right]^{0.5} = \sqrt{N}\sigma \le \sum_{i=1}^{N}|u_i - \mu| \Rightarrow \sigma \le \frac{1}{\sqrt{N}}\sum_{i=1}^{N}|u_i - \mu| = \sqrt{N}MAD$

Also, $\frac{1}{N}\sum_{i=1}^{N}|u_i - \mu| \ge \frac{1}{N}\sum_{i=1}^{N}|u_i - Median|$.
Thus, μ is the reference point that minimizes the sum of squared distances, while in the case of absolute distances, the reference point should be the median.

Sample variance and the (n−1 divisor

When estimating the variability among the population elements based on sample data, the choice of divisor in the definition of the sample variance becomes crucial. For an infinite population, dividing. $\sum_{i=1}^{n}(X_i - \bar{X})^2$by (n-1) yields an unbiased estimator$S^2$of the population variance,σ^2. This adjustment compensates for the use of the sample mean. $\bar{X}$ in place of the unknown population mean μ. However, it is important to note that while this correction ensures unbiasedness for this estimator, the corresponding sample standard deviation is. S remains biased.
Suppose that $X_1, X_2, \ldots, X_n$ is a random sample of size n taken from the population of interest. Then, the sample mean and variance are

$\bar{X} = \sum_{i=1}^{n} X_i \,/n$ & $S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2 \,/(n-1)\,, n \ge 2$.

Some Important Arguments About the (n – 1) Divisor:

1) When the underlying population is infinite, $(i.e.\ \ N = \infty)$This divisor is the only divisor that makes S^2an unbiased estimator of the population varianceσ^2; $E(S^2) = \sigma^2$. Now,

$$\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n}(X_i - \mu)^2 - n(\bar{X} - \mu)^2$$

Thus, dividing by n or by any other number $>n-1$ will give, on average, values of S^2 that are $< \sigma^2$. In other words, the$n-1$ is used because it is the only choice that makes S^2 an unbiased estimator of σ^2; $E(S^2) = \sigma^2$. But, what about S, the sample standard deviation. Actually since $S^2$is a convex function, using Jensen´s inequality, we have:
$E^2(S) < E(S^2) = \sigma^2 \Rightarrow E(S) < \sigma$, which means that S is biased.
In practice, we need to use S not S^2. Thus, why not choose a suitable divisor that makes S An unbiased estimator of σ ?

2) If $N < \infty$ (finite) then:

$\frac{N-1}{N(n-1)}E\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right) = \frac{N-1}{N(n-1)}\frac{N(n-1)}{N-1}\sigma^2 = \sigma^2$.

Thus, $\frac{N-1}{N}S^2$ is the unbiased estimator of σ^2 and not S^2. That makes this argument about the division by $(n-1)$ even improper.

3) Since, $\sum_{i=1}^{n}(x_i-\bar{x})^2$We have the freedom to choose $(n-1)$ of the n values and the last value is=− "Sum of the freely chosen values". That is a suitable reason for calling. $(n-1)$ "degrees of freedom"!
4) Clearly, to estimate variability, we need at least two values, $n \geq 2$; so the division by $n-1$ is a guard against using a sample of size 1 when we want to estimateσ^2. If $n=1$ then $S^2=\frac{0}{0}$, undefined. It is not ∞; ∞is the answer of a large number of students to the question of why n should not be $=1$?.
5) Assume that we have a random sample of n elements from a population with a mean$\mu$and varianceσ^2. The variability in the data set is measured by the standard deviation, which is naturally defined as the positive square root of the average squared deviations. Thus, using n as a divisor instead of $n-1$ gives:

$$S^{2^*}=\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2/n}$$

If $n=1$, then $S^*=0$, which as acceptable value of the variation when we have only one observation.
Now, when moving from descriptive to inferential statistics, we want to use the information in a suitably chosen representative sample to estimate the unknown parameters, μ &σ^2. $\bar{X}$ is always an unbiased estimator for finite μ. If the population is infinite, then $\frac{n}{n-1}S^{*2}$is an unbiased estimator ofσ^2. If the population is finite and of size N, then $\frac{n(N-1)}{(n-1)N}S^{*2}$is an unbiased estimator of σ^2.

The Variance of Two Subgroups(Strata)

Suppose a population is partitioned into two exhaustive and non-overlapping subgroups. The two subgroups are called two suitable strata if there are similarities within (the variance of each of them is significantly small compared to the variance of the population) and different among (the two groups have very different CVs). For example, when studying smoking habits, Males and Females are two suitable strata. A stratified random sample consists of two random sample one from each stratum. It is well known that the performance of stratified random sampling for estimating the population mean is better than that of simple random sampling(SRS) if the groups are suitable strata. For example: (2,5,7,8) and (2002,2004,2007,2008) are two suitable strata, but (1,100, 200) and (3, 200, 300) are not.
Let $u_1,u_2,\dots,u_{N_1}$be the values of the variable of interest for the elements of the first stratum and $w_1,w_2,\dots,w_{N_2}$be the values of the variable of interest for the elements of the second stratum. Let N be the population size. The size of the stratum is N_1 and the size of stratum 2 is N_2. Thus,

$\mu_1=\sum_{i=1}^{N_1}u_i/N_1,\ \mu_2=\sum_{i=1}^{N_2}w_i/N_2,$, and $\mu=(N_1\mu_1+N_2\mu_2)/N$

μ_1,μ_2,μ are respectively, the mean of the 1st stratum, the 2nd stratum, and the population mean. Also, the variances of strata 1 and 2 are:

$\sigma_1^2=\frac{1}{N_1}\sum_{i=1}^{N_1}u_i^2-\mu_1^2$ & $\sigma_2^2=\frac{1}{N_2}\sum_{i=1}^{N_2}w_i^2-\mu_2^2$.

Now, the variance of the population can be written as

$\sigma^2=\frac{1}{N}\left(\sum_{i=1}^{N_1}u_i^2+\sum_{i=1}^{N_2}w_i^2-N\mu^2\right)=\frac{1}{N}\sum_{i=1}^{N_1}u_i^2+\sum_{i=1}^{N_2}w_i^2-N\left(\frac{N_1\mu_1+N_2\mu_2}{N}\right)^2$. Thus, $\sigma^2=\frac{N_1\sigma_1^2+N_2\sigma_2^2}{N}+\frac{N_1N_2}{N^2}(\mu_1-\mu_2)^2$.

If $\mu_1=\mu_2$then $\sigma^2=\frac{N_1}{N}\sigma_1^2+\frac{N_2}{N}\sigma_2^2$, which is a convex combination of σ_1 & σ_2 .Thus, in this case, either σ_1 or σ_2 is larger than σ. Thus, the two groups are unsuitable as two strata if they have equal means.

Max(a,b) and Min(a,b)

For any two real numbers a and b, it can be shown easily that:

$Max(a,b)=\frac{1}{2}(a+b)+\frac{1}{2}|a-b|$ & $Min(a,b)=\frac{1}{2}(a+b)-\frac{1}{2}|a-b|$.

$\frac{1}{2}(a+b)$ is the average of the two numbers (μ) while, $\frac{1}{2}|a-b|$ is their standard deviation (σ). Thus,$Max(a,b)=\mu+\sigma$ and $Min(a,b)=\mu-\sigma$.

3. Concluding Remarks

In this paper, several less frequently known (hidden) properties of variance and related measures of variability have been examined. Attention was given to the maximum possible value of the variance, its relationship with the coefficient of variation, and its connection to the mean absolute deviation. The discussion also revisited the rationale behind the use of the $(n-1)$ as a denominator of the formula of the sample variance. These facts are not in the contents of most of the teaching textbooks. Presenting such topics in greater detail may help students develop a deeper and more coherent understanding of variability. Similar overlooked properties exist in many areas of statistics and mathematics, and exploring them offers promising directions for future work. One main query about the definition of mean absolute deviation is highlighted. Many other statistical concepts can be discussed in the future.

Acknowledgments

We thank the referee for the careful reading and inspection of the manuscript and for the very useful suggestions and comments, which led to a significant improvement of our article. Also, the efforts of the Editor members of the Editorial office are highly appreciated.

Funding and/or Conflicts of Interest/Competing Interests

The authors affirm that there is no conflict of interest regarding the publication of this article.

References

[1] AL-Saleh, M. F., "On the confusion about $(n-1)$-divisor of the standard deviation", Journal of Prob. and Stat. Science, vol. 5, 139-144, 2005.
[2] AL-Saleh, M. F., "A rich learning lesson using the Poisson distribution", Statistical Methodology, Vol. 4, no. 4, pp. 504-507, 2007. https://doi.org/10.1016/j.stamet.2007.01.005.
[3] Al-Saleh, M. F., Yousif, A. E., "Properties of the standard deviation that are rarely mentioned in classrooms", Austrian journal of statistics, vol. 3, pp. 193-202, 2009.
[4] Casella, G., Berger, R., "Statistical Inference", Wadsworth & Brooks/Cole Advanced Book & Software, California, USA, 1990.
[5] Balakrishnan, N., Balasubramanian, K., "Revisiting Sen's inequalities on order statistics", Statistics and prob. letters, vol. 78, pp. 616-621, 2007. https://doi.org/10.1016/j.spl.2007.09.023.
[6] Herbert, H. Mezbahur, R., "Revisiting the Digits of PI and Their Randomness", International Journal of Statistical Science (IJSS), vol. 4, pp. 13-24, 2005.
[7] Al-Saleh, M. F. and Maabreh, Arwa, "A Rich Learning Statistical Lesson in the Cauchy Distribution", Mutah Journal for Natural, Applied and Health Sciences, vol. 40, no. 1, pp. 13-32, 2025. https://doi.org/10.35682/mjnahs.v20i1.1192.
[8] Al-Saleh, M. F. and Jarrah, H., "Hidden Properties in the Main Sampling Techniques", International Journal of Advanced Statistics and Prob., vol. 12, no. 1, pp. 12-16, 2025. https://doi.org/10.14419/dp63md84.
[9] Heinbockel J. H., "Introduction to Calculus" Volume II, Old Dominion University, VA, USA, 2012.
[10] Khan, R., "A note on the generating function of a negative hyper-geometric distribution", Sankhya 56, 309-313, 1994.
[11] Bloch, D., "A note on the estimation of the location parameter of the Cauchy distribution", *Journal of the American Statistical Association*, 61(315), 852-855, 1966. https://doi.org/10.1080/01621459.1966.10480912.
[12] Copas J. B., "On the unimodality of the likelihood for the Cauchy distribution," Biometrika, 62, no.3, 701-704, 1975. DOI: 10.1093/biomet/62.3.701. https://doi.org/10.1093/biomet/62.3.701.
[13] Rothenberg T. J., Fisher F. M., Tilanus C. B., " A note on estimation from a Cauchy sample," Journal of the American Statistical Association, vol. 59, no. 306, pp.460-463, 1964. https://doi.org/10.1080/01621459.1964.10482170.
[14] Azmay N., Rosli R., Maat S., Mahmud M. "Reasoning and Statistical Thinking: A Systematic Literature Review", International journal of academic research in progressive education and development, Vol. 1 2 , No. 2, 2023, E-ISSN: 2226-6348 © 2023.
[15] Schoen, C. , Christopher, B ., Alexandra B., Tim Jacobbe C., Lanrong, d. " Improving the Teaching and Learning of Statistics", ILearning and Instruction, Vol. 95, 2025, 102018. https://doi.org/10.1016/j.learninstruc.2024.102018.